JOURNAL OF BACTERIOLOGY, Dec. 2008, p. 7838–7846 Vol. 190, No. 23
0021-9193/08/$08.00+0 doi:10.1128/JB.00827-08

Frequent Homologous Recombination Events in *Mycobacterium tuberculosis* PE/PPE Multigene Families: Potential Role in Antigenic Variability▽†‡

Anis Karboul,[1] Alberto Mazza,[2] Nicolaas C. Gey van Pittius,[3] John L. Ho,[4] Roland Brousseau,[2] and Helmi Mardassi[1]*

Unit of Typing and Genetics of Mycobacteria, Institut Pasteur de Tunis, Tunis, Tunisia[1]; Biotechnology Research Institute of Montreal, 6100 Royalmount, Montreal, Canada[2]; DST/NRF Centre of Excellence in Biomedical Tuberculosis Research, MRC Centre for Molecular and Cellular Biology, Department of Biomedical Sciences, Faculty of Health Sciences, Stellenbosch University, Tygerberg, South Africa[3]; and Division of International Medicine and Infectious Diseases, Weill Medical College of Cornell University, New York, New York[4]

Received 13 June 2008/Accepted 20 September 2008

The PE and PPE (PE/PPE) multigene families of *Mycobacterium tuberculosis* are particularly GC-rich and share extensive homologous repetitive sequences. We hypothesized that they may undergo homologous recombination events, a mechanism rarely described in the natural evolution of mycobacteria. To test our hypothesis, we developed a specific oligonucleotide-based microarray targeting nearly all of the PE/PPE genes, aimed at detecting signals for homologous recombination. Such a microarray has never before been reported due to the multiplicity and highly repetitive and homologous nature of these sequences. Application of the microarray to a collection of *M. tuberculosis* clinical isolates ($n = 33$) representing prevalent spoligotype strain families in Tunisia allowed successful detection of six deleted genomic regions involving a total of two PE and seven PPE genes. Some of these deleted genes are known to be immunodominant or involved in virulence. The four precisely determined deletions were flanked by 400- to 500-bp stretches of nearly identical sequences lying mainly at the conserved N-terminal region of the PE/PPE genes. These highly homologous sequences thus serve as substrates to mediate both intergenic and intragenic homologous recombination events, indicating an important function in generating strain variation. Importantly, all recombination events yielded a new in-frame fusion chimeric gene. Hence, homologous recombination within and between PE/PPE genes likely increased their antigenic variability, which may have profound implications in pathogenicity and/or host adaptation. The finding of high prevalence (~45% and ~58%) for at least two of the genomic deletions suggests that they likely confer advantageous biological attributes.

The mycobacterial genome contains two large gene families encoding acidic or asparagine- or glycine-rich proteins, referred to as PE and PPE proteins (14). In *Mycobacterium tuberculosis* strain H37Rv, PE ($n = 99$) and PPE ($n = 68$) genes were found to be scattered throughout the whole genome, accounting for approximately 10% of its coding capacity. Both families are characterized by their high GC content and extensive repetitive homologous sequences. Much of these characteristics is associated with the two major subgroups, PE_PGRS ($n = 61$) and PPE-MPTR ($n = 23$) (9, 14, 23). Expansion of the mycobacterially restricted PE/PPE gene families, most likely through gene duplication, appears to have accompanied the evolutionary history of tubercle bacilli, being linked to the duplication of the ESAT-6 gene clusters (23).

PE/PPE genes are strongly suspected to be involved in several aspects of host-pathogen interactions, such as antigenic variability, virulence, and persistence of the bacillus (4, 8, 13, 15–17, 40, 41, 43, 45). Their surface-exposed domains are also involved in the shaping of the bacterial cell structure (16). Some contiguous PE and PPE genes were shown to be capable of forming heterodimers which are secreted and thought to play a role in signal transduction (44, 49).

Considerable sequence polymorphism has been reported for several members among the 168 PE/PPE genes (14, 19, 21, 22, 25, 26–28, 34, 46, 47). These multigene families thus appear to be the major contributors to the genetic diversity of *M. tuberculosis*, an organism assumed to have a relatively invariable genome (14). Aside from single nucleotide polymorphisms, a frequent number of deletions have been reported within PE/PPE genes. Since these multigene families are enriched in homologous sequences and occur mostly as a cluster of several contiguous members throughout the genome, they might represent an optimal environment for homologous recombination. Hence, deletions between these homologous sequences might be a potent indicator of the occurrence of recombination in *M. tuberculosis*.

Detection of deletions associated with PE/PPE genes is difficult, as their particular structure is incompatible with hybridization-based strategies such as microarrays. The difficulties encountered in interpreting hybridization signals associated with PE/PPE genes have been highlighted in many studies (29, 38, 48). In addition, validating alterations within PE/PPE genes by PCR amplification is inherently difficult due to their high G+C content and richness in repetitive sequences. In fact, deletions in PE/PPE genes have mostly been inferred from comparative genomics (19, 22, 25) or fortuitously identified as part of a larger deletion event (6, 21, 31, 36, 38, 48).

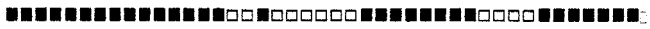
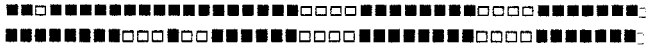
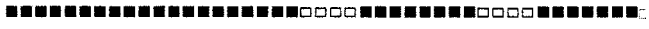
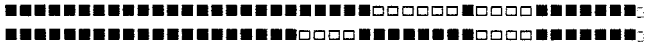
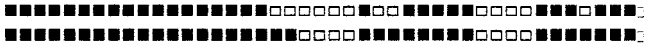
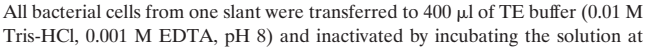
* Corresponding author. Mailing address: Institut Pasteur de Tunis, 13 Place Pasteur, B.P. 74, 1002 Tunis-Belvédère, Tunis, Tunisia. Phone: (216) 71 843 755. Fax: (216) 71 791 833. E-mail: helmi.merdassi@pasteur.rns.tn.

† Supplemental material for this article may be found at http://jb.asm.org/.

‡ This paper is dedicated to the memory of Roland Brousseau.

▽ Published ahead of print on 26 September 2008.

TABLE 1. *M. tuberculosis* clinical isolates from Tunisia

Strain	Geographical origin	Spoligotype family[a]	Spoligotype pattern
Tn1	Tunis	Orphan (T3-like)	
Tn2	Menzel Bourguiba	LAM9	
Tn3	Tunis	H3	
Tn4	Tunis	Orphan (T3-like)	
Tn5	Tunis	Orphan (LAM9-like)	
Tn6	Tunis	Orphan (LAM9-like)	
Tn7	Tunis	Orphan (LAM1)	
Tn8	Zaghouan	Orphan (LAM3-like)	
Tn9	Zaghouan	LAM9	
Tn10	Menzel Bourguiba	H1	
Tn11	Menzel Bourguiba	H1	
Tn12	Tunis	LAM9	
Tn13	Tunis	H1	
Tn14	Menzel Bourguiba	T1	
Tn15	Menzel Bourguiba	Orphan (T4-like)	
Tn16	Menzel Bourguiba	Orphan (H1-like)	
Tn17	Tunis	H3	
Tn18	Menzel Bourguiba	H3	
Tn19	Tunis	H1	
Tn20	Menzel Bourguiba	LAM9	
Tn21	Menzel Bourguiba	Orphan (T1-like)	
Tn22	Tunis	LAM9	
Tn23	Menzel Bourguiba	H3	
Tn24	Tunis	Orphan (LAM9-like)	
Tn25	Tunis	LAM9	
Tn26	Menzel Bourguiba	H1	
Tn27	Menzel Bourguiba	Orphan (LAM9-like)	
Tn28	Menzel Bourguiba	T1	
Tn29	Menzel Bourguiba	H3	
Tn30	Menzel Bourguiba	H3	
Tn31	Menzel Bourguiba	H3	
Tn32	Menzel Bourguiba	H3	
Tn33	Menzel Bourguiba	H3	

[a] Genotypes were determined according to SpolDB4. Types in parentheses indicate spoligotype families determined by Spotclust analysis.

To our knowledge, a DNA microarray specifically designed to target deletions within PE/PPE genes and whose specificity was addressed has never been reported. In this study, we describe an optimized oligonucleotide-based microarray protocol aimed at detecting large sequence polymorphisms (LSPs) within PE/PPE genes, encompassing nearly the entire collection of known members. The microarray-derived signals provided reasonable specificity and sensitivity levels and allowed the mapping of numerous PE/PPE-associated deletions that were confirmed by nucleotide sequencing. Overall, the findings provide evidence for the frequent occurrence of homologous recombination within and between PE/PPE genes, thereby significantly contributing to their own diversification as a putative source of antigenic variability and host adaptation.

MATERIALS AND METHODS

Mycobacterial strains and DNA extraction. For optimization of the 50-mer oligonucleotide-based microarray, we used a *Mycobacterium bovis* bacillus Calmette Guérin (BCG) isolate recovered from the lymph node of a Tunisian patient as well as two sequenced *M. tuberculosis* strains, H37Rv and CDC1551, and a Brazilian clinical *M. tuberculosis* strain (318/02). The latter harbors a large genomic deletion (RD^{Rio}) involving two PPE genes (PPE55 and PPE56) (31). Thirty-three Tunisian *M. tuberculosis* isolates representative of the major circulating genotypes (37), originating from three regions in the north of the country, were used to detect LSPs in PE/PPE genes and to assess the sensitivity and specificity of the microarray. The list and characteristics of the *M. tuberculosis* isolates are presented in Table 1. Mycobacterial strains were cultured in slanted culture tubes on Löwenstein-Jensen medium according to standard protocols. All bacterial cells from one slant were transferred to 400 μl of TE buffer (0.01 M Tris-HCl, 0.001 M EDTA, pH 8) and inactivated by incubating the solution at 95°C for 30 min. Whole genomic DNA was then extracted by using a standard phenol-chloroform extraction procedure (50).

Oligonucleotides. A preliminary list of 194 50-mer oligonucleotides was selected from full-length PE/PPE gene sequences of *Mycobacterium* spp. The oligonucleotides were selected using the public domain software OligoPicker (53). All oligonucleotides were selected from an input file made from the known list of PE/PPE genes present in *M. tuberculosis*, *M. bovis*, and *Mycobacterium avium* subsp. *paratuberculosis*. The parameters used were a maximum continuous match to nontarget sequences of 16 nucleotides and a range of melting temperatures of 10°C. Oligonucleotides were tested by BLAST analysis through the TubercuList (http://genolist.pasteur.fr/TubercuList/index.html) or BoviList (http://genolist.pasteur.fr/BoviList/index.html) site, and those that showed more than 50% sequence identity with another genomic location were discarded. Among the latter selected oligonucleotides, those that failed to discriminate between the test strains H37Rv, CDC1551, and *M. bovis* were eliminated. The finalized list contained 172 oligoprobes, whose nucleotide sequences and targeted genes are available upon request.

Construction of the microarray. The selected 50-base oligonucleotides were purchased from a commercial supplier (IDT, Coralville, IA). They were printed on Corning Ultra GAPS slides (Corning Canada, Whitby, Ontario, Canada), using a commercially available Bio-Rad Versarray Pro robotics array printer at the microarray laboratory of the Biotechnology Research Institute, National Research Council of Canada, Montreal, Quebec, Canada.

Labeling of genomic DNA. DNA was labeled using a random primer linear amplification method (BioPrime DNA labeling system; Invitrogen), replacing biotinylated dCTP incorporation with Cy5 fluorophore-tagged dCTP incorporation. Frequency of incorporation was targeted at 10 to 12 units/10^3 bases.

Hybridization of labeled genomic DNA. Two 50-mer microarrays constructed as described above were probed with mycobacterial labeled genomic DNA. Hybridization under highly stringent conditions was effected for the second array

in DIG Easy Hyb buffer supplemented with formamide to 20%, salmon sperm single-stranded DNA (final concentration, 1.0 μg/μl), and purified genomic DNA from GC-rich *Methylbacterium extorquens* (final concentration, 0.1 μg/μl) at 50°C overnight with a Slidebooster hybridization station (Advalvtixs SB800; The GelCompany, San Francisco, CA). Microarray scans were performed using a Canberra-Packard ScanArray Lite laser. Power output and photomultiplier gain of the microarray scanner were adjusted according to the signal available. For both hybridization conditions, the scan was taken with the same scanner and the same laser and photomultiplier tube settings. Numerically extracted hybridization data (Imagene8.0) were converted to values representing the change over background per spot. Binary scores were set with a threshold of 4× over background, whereby scores of ≥4.0 were set at 1 and scores of <4.0 were set at 0.

Phylogenetic reconstruction. The microarray data were converted to binary scores (positive hybridization signal = 1 and absence of hybridization = 0) and analyzed by GeneCluster 3.0 (Eisn Lab [freeware]), using the hierarchical clustering on genes (similarity metric/correlation [uncentered]) with complete linkage method. Resulting clustered data were visualized using JavaTreeView.

Confirmation of deletions by PCR. Confirmatory analyses of the presence or absence of deletions were performed by PCR amplification using primer pairs placed at different distances upstream and downstream of the probe. Functional primer pairs targeting the variable regions are listed in Table S1 in the supplemental material. The PCR amplification reaction mixture contained 2 μl of template genomic DNA (about 20 ng), 5 μl of 10× buffer (Qiagen), 5 μl of dimethyl sulfoxide, 1 μl of 10 mM nucleotide mix (Amersham Biosciences), 2 μl of each primer (20 μM stock concentration), 0.25 μl (1.25 U) of HotStar *Taq* DNA polymerase (Qiagen), and sterile nuclease-free water (Amersham Biosciences) to a 50-μl total reaction volume. Cycling was carried out in a PXE 0.2 thermal cycler (Thermo Electron Corporation), with an initial denaturation step of 10 min at 96°C, followed by 35 cycles consisting of 1 min at 95°C, 1 min at 60°C, and 2 min at 72°C. The amplification ended with a final elongation step of 7 min at 72°C.

Sequencing of deletion junction regions. Amplicons with a reduced size, suggestive of a true deletion, as well as those with a typical mobility pattern were subjected to nucleotide sequencing to determine, to the base pair, the junction regions of the deletion and to check for the presence or absence of mutations. This was done for all the tested isolates in order to confirm whether shared deletions were identical.

Amplicons were subjected to sequencing after treatment with exonuclease I (Amersham Biosciences) and shrimp alkaline phosphatase (Amersham Biosciences). The reaction mixture consisted of 1.5 μl of BigDye Terminator cycle sequencing reagents, 4 μl of BigDye Terminator cycle sequencing buffer, 1 μl of primers (20 μM), and sufficient ultrapure distilled DNase- and RNase-free water (Gibco/Invitrogen) to make a 20-μl reaction mixture. Cycle sequencing was performed using a PXE 0.2 thermal cycler (Thermo Electron Corporation) programmed for 25 cycles at 96°C for 10 s, 50°C for 5 s, and 60°C for 4 min. The template DNA was ethanol precipitated, washed, and subjected to automated sequencing on an ABI Prism 3130 genetic analyzer according to the manufacturer's protocol.

RESULTS

Design and optimization of DNA microarray protocol. All PPE genes, except for two ($n = 66$), and all PE genes, except for five ($n = 94$), were represented in the microarray. The two PPE and five PE genes for which specific probes (according to our selection criteria) could not be identified are Rv3022c (PPE48), Rv3125c (PPE49), Rv2408 (PE24), Rv2126c (PE_PGRS37), Rv0978c (PE_PGRS17), Rv2591 (PE_PGRS44), and Rv3018A (PE27A). There was no significant variability within these individual genes which allowed the identification of any region large enough for design of a specific probe. Because each PE/PPE gene is represented by a single oligonucleotide, our analysis could thus mainly detect LSPs.

Hybridization of the microarray under normal conditions (42°C in DIG Easy Hyb buffer overnight in a Slidebooster hybridization station) resulted in a large number of cross-hybridizations (data not shown). This was expected given the high GC content and repetitiveness of PE/PPE sequences. Hence, we varied the hybridization parameters and found that optimal conditions consisted of a hybridization temperature of 50°C overnight (in DIG Easy Hyb buffer in a Slidebooster staion) in the presence of formamide, salmon sperm single-stranded DNA, and purified GC-rich genomic DNA from *Methylobacterium extorquens*. Under such highly stringent conditions, all shared and strain-specific PE/PPE genes were efficiently detected, and all of the known LSPs in PE/PPE genes in H37Rv (genes Mb3184c and Mb2376c, according to the *M. bovis* annotation), CDC5151 (the PPE54, PPE57, and PPE58 genes, according to the *M. tuberculosis* H37Rv annotation), and the Brazilian RD^{Rio}-containing deletion strain (genes PPE55 and PPE56, according to the *M. tuberculosis* H37Rv annotation) were unambiguously revealed (Fig. 1). Confirmation of the efficiency of the microarray was provided by the BCG strain, which showed no hybridization signal with 13 deleted or truncated PE/PPE genes (PE32, PE35, PE_PGRS35, PPE38, PPE39, PPE40, PPE54, PPE57, PPE58, PPE65, PPE66, PPE67, and PPE68) (6, 10, 11, 24).

Genomic interrogation of a set of clinical *M. tuberculosis* isolates, using the PE/PPE-specific oligonucleotide-based microarray. Using the aforementioned optimized hybridization conditions, we next searched for the occurrence of LSPs in PE/PPE genes throughout a set of 33 clinical isolates representative of the prevalent *M. tuberculosis* spoligotype strain families circulating in Tunisia (12, 37) (Table 1). Overall, based on the hybridization pattern of the microarray, 23 probes were considered informative, since each yielded at least one instance where no hybridization signal was detected with one strain from the *M. tuberculosis* clinical isolate collection (33 isolates). Among the 759 signals emitted by these 23 informative probes, there were 747 correct results (based on PCR as the gold standard) versus 12 false or ambiguous results (8 were false-negative results, 1 was a false-positive result, and 3 were ambiguous) (Fig. 2). Using a conservative approach of counting the three ambiguous results as false, the hybridization results showed a specificity of 96.2% and a sensitivity of 98.7%.

Six LSPs, designated LSP^{Tun1} to LSP^{Tun6}, were identified (Table 2). We were successful in characterizing four of them (LSP^{Tun2}, LSP^{Tun4}, LSP^{Tun5}, and LSP^{Tun6}), and they were PCR amplified for sequence analysis (see below). The junction regions of the two remaining deletions (LSP^{Tun1} and LSP^{Tun3}) could not be PCR amplified, even though several attempts were made using different primer pairs. Since these LSPs were unambiguously shared by several isolates (Fig. 2 and Table 3), they were considered true deletions. Moreover, deletions that map to these regions have been reported previously (36, 48). The presence or absence of LSPs was further confirmed by PCR amplification. Overall, the identified deletions comprised two PE genes (PE18 and the PE_PGRS gene *wag-22*) and seven PPE genes (PPE25, PPE26, CDC1551 PPE MT3248, PPE57, PPE58, PPE54, and PPE47).

Mapping the deletion boundaries to the base pair. As mentioned above, only four of the six deletions could be confirmed by PCR amplification of the boundary regions. Primers flanking these deleted regions did indeed yield smaller amplicons than the expected size for all isolates bearing these deletions (data not shown). These were subjected to nucleotide sequence analysis to determine the deletion junctions to the base pair. The deletions varied in size from 2,128 bp to 5,283 bp

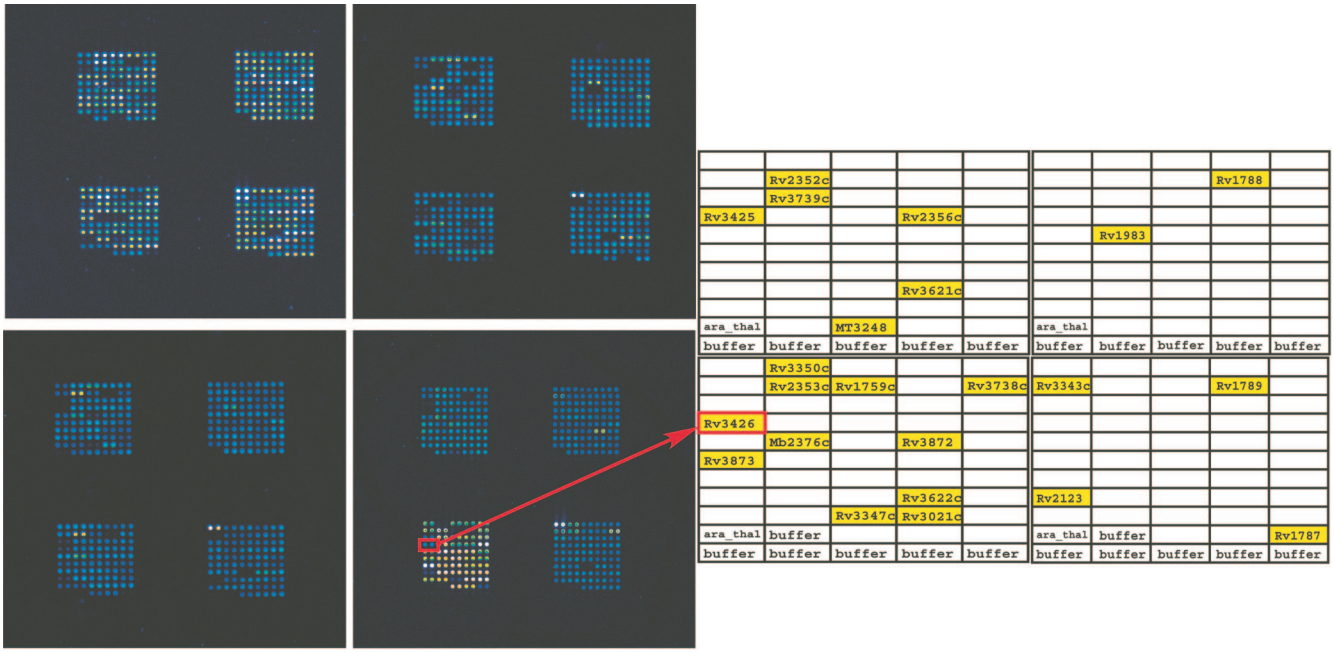


FIG. 1. Hybridization results for the four reference strains (H37Rv, 1629/05 [BCG], 318/02 [RD^{Rio}], and CDC1551) used in this study. The schematic grid on the right represents the array positions of the 23 informative probes (shown in yellow), seen as paired spots on the actual scanned images (left), as well as the positions of the negative controls (*Arabidopsis* probes and buffer spots). In the hybridization grid for the RD^{Rio}-containing strain 318/02, arrow A points to an ambiguous array signal (just below the signal threshold for positive results).

(Table 2). Sequence analysis showed that for each deletion the junction sequences were identical in all isolates in which it occurred, suggesting a common ancestor.

LSP^{Tun2} is a newly uncovered genomic location that has never been reported to have lost any genetic material. In contrast, the deletions mapped to LSP^{Tun4}, LSP^{Tun5}, and LSP^{Tun6} overlapped previously described deletions by between 91% and 95%. However, these three LSP loci were dissimilar at the chimeric gene junctions from those previously described, indicating the occurrence of multiple independent genetic events in their creation, and as such, these loci may be hot spots for deletion. As shown in Fig. 3, three of the LSPs (LSP^{Tun2}, LSP^{Tun4}, and LSP^{Tun6}) involved two or more PE/PPE members, even resulting in the in-frame fusion of distant genes (e.g., in the case of LSP^{Tun2} and LSP^{Tun6}). Interestingly, all deletions with confirmed boundaries were found to be flanked, in their original nondeleted wild-type form, by identical or nearly identical sequences of ~400 to ~500 bp in length (Fig. 3; see Fig. S1 in the supplemental material). In three of the four cases, these homologous flanking regions corresponded nearly exactly to the complete conserved 540-nucleotide N-terminal regions of the PPE genes, indicating an important function for these conserved regions in generating strain variation. This is in contrast to previous suggestions that variation was driven primarily by the C-terminal repetitive regions of these genes. In each deleted region, a single copy of the two identical sequences was found to be present. In addition, all of the deletions maintained the sequence continuity of the two parental genes, leading to a fully functional new gene. These characteristics (deletion through identical flanking sequences and maintenance of sequence integrity) are reminiscent of typical homologous recombination events. Moreover, we could not identify any insertional sequence in the vicinity of the deleted regions, indicating that transposons were not involved in these specific deletion events.

Frequency and distribution of LSPs in the Tunisian *M. tuberculosis* collection. The LSP^{Tun3} and LSP^{Tun4} deletions were the most frequently encountered deletions, occurring in 57.4% and 45.4% of isolates, respectively (Table 3). LSP^{Tun3} was shared by most of the strain families, including Haarlem, LAM, and T, but was more associated with Haarlem strains (91.6%). In contrast, LSP^{Tun4}, the second most frequent deletion shared by these three strain families, was more frequently linked to the LAM family (66.7%). Notably, the reference *M. tuberculosis* strain H37Rv also contains a deletion that maps to LSP^{Tun4} (19). The LSP^{Tun6} deletion, which maps to the RD6 region, was present in 30% of the strains and tended to occur in LAM and T genotypes. Finally, LSP^{Tun2} and LSP^{Tun5} deletions were the least frequent deletions, being associated with ~12% and ~6.1%, respectively, of the data set (Table 3). LSP^{Tun2} was restricted solely to the LAM9 genotype.

DISCUSSION

Although it has been demonstrated clearly in vitro that slow-growing mycobacteria are not refractory to homologous recombination (2, 3, 5, 7), the occurrence of such events under natural conditions for the *Mycobacterium* genus as a whole has rarely been documented (30, 31). We previously provided evidence for a naturally occurring homologous recombination event, namely, gene conversion, involving a PE_PGRS duplicated gene pair (28). This finding prompted us to further explore the entire PE/PPE gene families in the hope of detecting additional evidence for homologous recombination specific

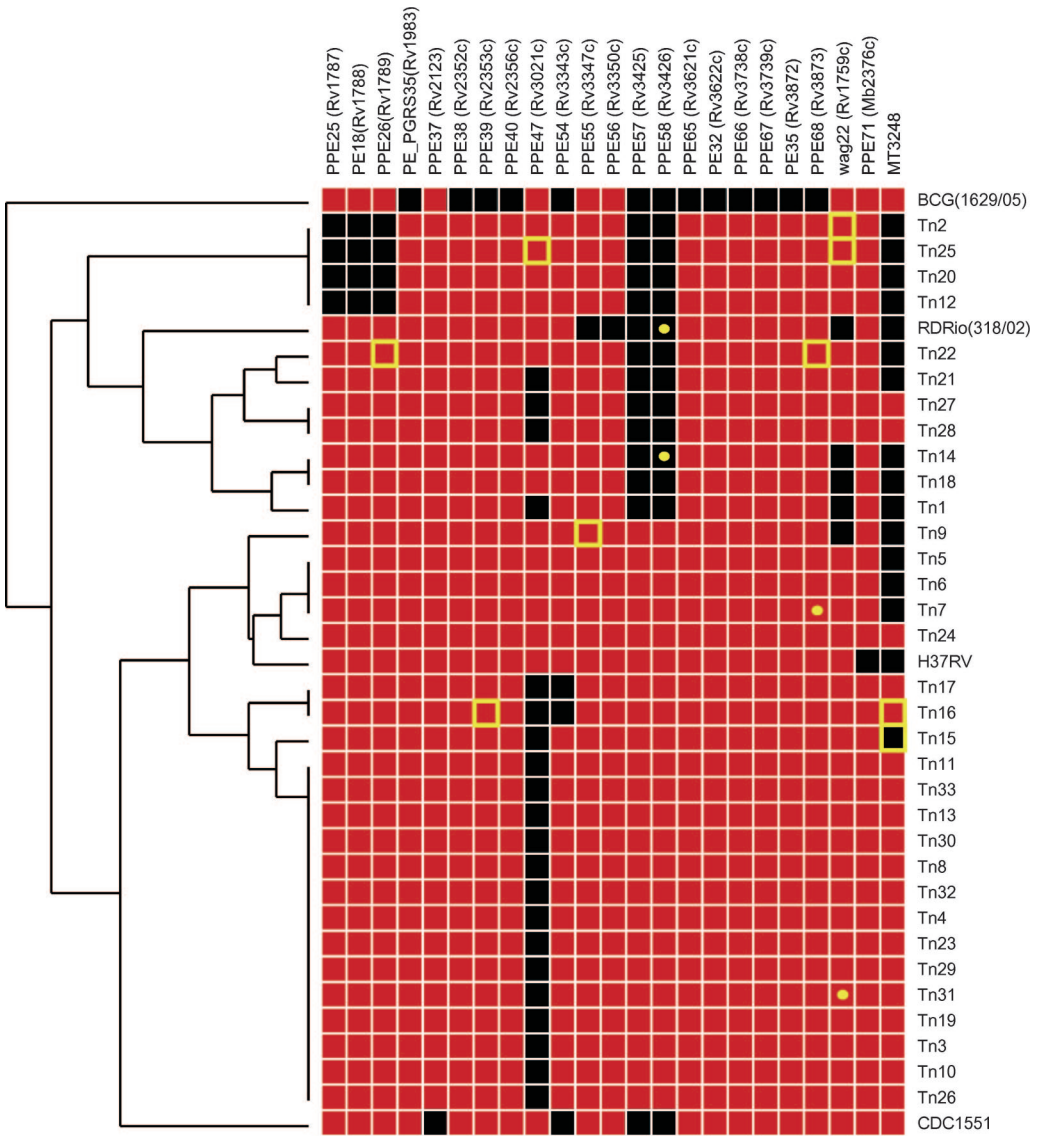


FIG. 2. Dendrogram and hybridization results for *M. tuberculosis* strains from Tunisia compared to those for reference strains BCG, CDC 1551, H37Rv, and 318/02 (RD^{Rio}). Red squares correspond to positive hybridization results, and black squares correspond to negative hybridization results. Squares framed in yellow represent divergent results between PCR and hybridization; in these cases, the PCR result was deemed the correct one and used to build the dendrogram. Squares containing a small yellow circle represent ambiguous hybridization results; in these cases, the PCR result was also deemed correct. The ambiguous hybridization result for PPE58 in RD^{Rio} was not checked by PCR but was considered negative given that the result for PPE57 in RD^{Rio} was clearly negative and that these two genes appear to be highly correlated in our data.

to these families. For this purpose, we developed a microarray targeting PE/PPE genes and optimized the hybridization protocol in order to specifically detect LSP events between PE/PPE sequences. Complementary sequence analysis would provide evidence indicative of homologous recombination.

Overall, the results indicated that the developed microarray, combined with the optimized hybridization protocol, offered very high levels of sensitivity and specificity in detecting LSPs. Application of the microarray to a collection of clinical *M. tuberculosis* isolates representative of prevalent spoligotype strain families identified six LSPs involving PE/PPE genes. Four of these were successfully characterized fully by sequence analysis. All of these characterized deletions start and end within PE/PPE sequences, thus confirming the efficacy of the microarray to solely target PE/PPE genes. All but one PE/PPE deletion occurred within regions previously reported to be polymorphic (6, 19, 21, 29, 38, 48), consistent with previous studies reporting that genomic deletions in *M. tuberculosis* tend to aggregate to hot spots that appear to be more vulnerable to deletion (48). Importantly, further scrutiny of the nucleotide sequence environment upstream and downstream of the four characterized deletions revealed that they were all immediately flanked by a stretch of identical (or nearly identical) sequences. This finding, along with the maintenance of sequence integrity and the absence of mobile elements, confirms our supposition that PE/PPE genes undergo natural recombination between their homologous sequences and supports our previous finding relating to the occurrence of gene conversion within very ho-

TABLE 2. Six deletion events in PE/PPE gene loci, indicated as LSPs, identified in *M. tuberculosis* clinical isolates from Tunisia ($n = 33$)

Deletion	Coordinate[a]		Size of deletion (bp)	Deleted gene(s)	Putative function(s)	Observations
	Start	End				
LSP^{Tun1}	ND	ND	ND	Rv1759c	PE_PGRS gene with fibronectin binding activity, referred to as *wag-22* (18)	A deletion involving *wag-22* was previously reported by Tsolaki et al. (48); Wag22 was shown to elicit a humoral immune response
LSP^{Tun2}	2025301	2028426	3,125	Rv1787, Rv1788, Rv1789, Rv1790[b]	PPE25 gene, implicated in virulence of *M. avium* (33); PE18 gene, unknown function; PPE26 gene, unknown function; and PPE27 gene, unknown function	This deletion has never been reported hitherto; it maps to the so-called ESX-5 cluster, whose orthologue in *M. marinum* was shown to be implicated in secretion of a heterologous PPE gene (1); the orthologue of PPE25 in *M. avium* was shown to be associated with the ability to grow in macrophages (33)
LSP^{Tun3}	ND	ND	ND	Rv3021c	PPE47 gene, unknown function	A deletion mapping to this region was previously reported by Marmiesse et al. (36); they noticed the presence of identical sequences flanking the deletion
LSP^{Tun4}	3524367	3526495[a]	2,128	MT3247,[c] MT3248	PPE genes, unknown function	Overlaps by 91.6% of its size a deletion previously described for CDC1551 (19)
LSP^{Tun5}	3730814	3736097	5,283	Rv3343c	PPE54 gene; unknown function	Overlaps by 95.5% of its size at the 3′ end a deletion previously described for CDC1551 (19)
LSP^{Tun6}	3841871	3846797	4,926	Rv3425, Rv3426, Rv3427c, Rv3428c, Rv3429[d]	PPE57 gene, unknown function but immunodominant; PPE58 gene, unknown function; gene involved in the transposition of IS*1532*; gene involved in the transposition of IS*1532*; and PPE59 gene, unknown function	Maps to the RD6 deletion (10), where it overlaps by 94.05% of its size; it eliminates PPE57, which is known to induce strong B- and T-cell responses (52, 54)

[a] Coordinates according to *M. tuberculosis* strain CDC1551.
[b] Partially deleted; merged in frame to Rv1787 (PPE25).
[c] Partially deleted; merged in frame to MT3248.
[d] Partially deleted; merged in frame to Rv3425 (PPE57).

mologous and adjacent PE genes (28). Despite the limited number of isolates analyzed and the fact that they are from a geographically confined area, the data suggest that PE/PPE-associated homologous recombination in *M. tuberculosis* likely occurs more frequently under natural conditions. These findings contrast with the view that homologous recombination is rare in mycobacteria.

PE/PPE protein superfamilies are surface exposed and represent putative antigens in mycobacterial virulence and the host immune response (4, 14, 40). Consequently, the sequence variability mediated by naturally occurring homologous recombination events, as demonstrated in this study, should be considered an effective mechanism contributing to the creation of new antigenic diversity for the host immune response and of new determinants for pathogen virulence. In our prior study, we provided the most direct evidence that in modern-day *M. tuberculosis* strains, PE genes exchange polymorphic sites flanked by homologous sequences through gene conversion (without a loss of intervening sequences), which ultimately contributes to increasing their diversity and hence their antigenic properties (28). The current study expands upon this finding and provides examples of inter- and intra-PE/PPE-gene homologous recombination events removing large sequences but, importantly, leading to in-frame gene fusions that now encode putative "new" protein antigens. The relatively high prevalence of LSP^{Tun3}, LSP^{Tun4}, and LSP^{Tun6} strains in a clonal expansion of *M. tuberculosis* in Tunisia argues that the removal of the intervening genes certainly did not reduce transmissibility and host pathogenicity. It is likely that other, very homologous PE/PPE genes could compensate for the deleted genes. It is also tempting to speculate that the loss of immunodominant PE/PPE genes may increase the bacillus's

TABLE 3. Distribution of LSP^{Tun} deletions in *M. tuberculosis* isolates from Tunisia

Spoligotype	Isolate	Presence of deletion[a]					
		LSP^{Tun1}	LSP^{Tun2}	LSP^{Tun3}	LSP^{Tun4}	LSP^{Tun5}	LSP^{Tun6}
H3	Tn29	−	−	+	−	−	−
	Tn30	−	−	+	−	−	−
	Tn31	−	−	+	−	−	−
	Tn32	−	−	+	−	−	−
	Tn33	−	−	+	−	−	−
	Tn17	−	−	+	−	+	−
	Tn3	−	−	+	−	−	−
	Tn23	−	−	+	−	−	−
H1	Tn19	−	−	+	−	−	−
	Tn10	−	−	+	−	−	−
	Tn11	−	−	+	−	−	−
	Tn26	−	−	+	−	−	−
H1-like	Tn16	−	−	+	−	+	−
LAM9	Tn2	−	+	−	+	−	+
	Tn20	−	+	−	+	−	+
	Tn25	−	+	−	+	−	+
	Tn12	−	+	−	+	−	−
	Tn22	−	−	−	+	−	+
	Tn9	+	−	−	+	−	−
LAM4	Tn18	+	−	−	+	−	+
LAM9-like	Tn5	−	−	−	+	−	−
	Tn6	−	−	−	+	−	−
	Tn27	−	−	−	−	−	+
	Tn24	−	−	−	−	−	−
LAM1-like	Tn7	−	−	−	+	−	−
LAM3-like	Tn8	−	−	−	+	−	−
T1	Tn28	−	−	+	−	−	+
	Tn13	−	−	+	−	−	−
	Tn14	+	−	−	+	−	−
T1-like	Tn21	−	−	+	+	−	+
T3-like	Tn1	+	−	+	+	−	+
	Tn4	−	−	+	−	−	−
T4-like	Tn15	−	−	+	+	−	−

[a] The frequencies of the deletions were 12.1, 12.1, 57.5, 45.4, 6, and 27.7% for LSP^{Tun1}, LSP^{Tun2}, LSP^{Tun3}, LSP^{Tun4}, LSP^{Tun5}, and LSP^{Tun6}, respectively.

capability to circumvent the host immune system, thus contributing to its success.

Intergenic homologous recombination, like in the case of LSP^{Tun2}, LSP^{Tun4}, and LSP^{Tun6}, not only results in the elimination of the intervening genes and the creation of new recombinant genes but could potentially introduce a dramatic change in the expression pattern of PE/PPE genes, for example, through loss or gain of regulatory elements. Indeed, recombination between distantly positioned PE/PPE genes could bring newly recombined sequences under the control of a heterologous promoter, thus increasing the heterogeneity of expression patterns between different isolates. Such an event could explain the heterogeneity in PE/PPE expression between clinical isolates previously reported by cDNA microarray (20, 51). LSP^{Tun5} involved a single gene, PPE54, which is one of the largest PPE genes. The sequence of this gene reveals three identical regions, leaving open the possibility for the occurrence of combinatorial recombination to increase antigenic variability.

Even though deletion in PE/PPE genes seems counterintuitive on first evaluation, the data support the hypothesis that elimination or modification of PE/PPE genes may indeed be a mechanism to enhance the pathogenicity of tubercle bacilli. The relatively increased disease severity of *M. tuberculosis* strain Erdman compared to that of H37Rv in the rabbit model was attributed mainly to gene differences in the RD6 region, which contains several PPE genes, including the immunodominant PPE57 gene (35, 52, 54). The recent characterization of the RD^{Rio} sublineage of the Latin American-Mediterranean strain family further bolsters the argument that strains bearing PE/PPE deletions likely confer successful biological attributes to their host interactions (24, 31). Importantly, patients harboring RD^{Rio} strains appeared to manifest a different disease spectrum that included a higher rate of cavitary disease (32), a form of tuberculosis that is associated with higher sputum bacillary loads (39) and thus a greater propensity to be transmitted to new hosts.

LSP^{Tun2}, which maps to the secretion locus ESX-5 (1), in which the PPE25 gene is deleted, is another interesting dele-

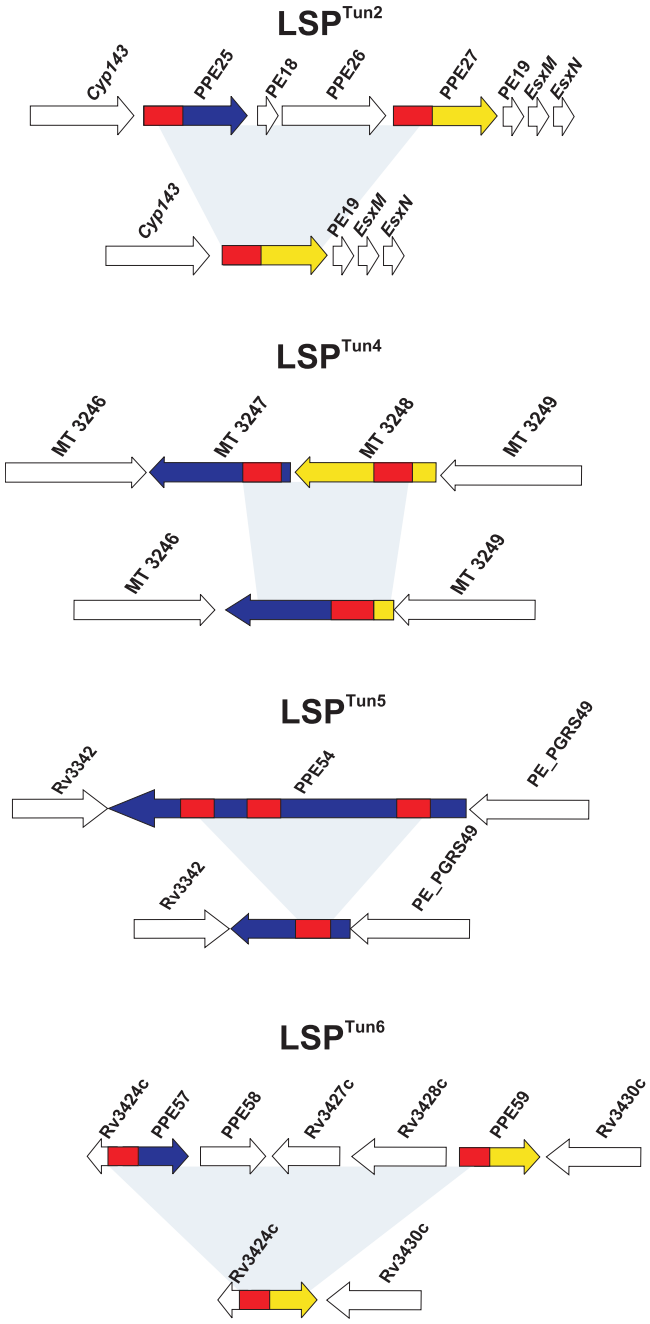


FIG. 3. Schematic representation of the genomic loci corresponding to LSP^{Tun2}, LSP^{Tun4}, LSP^{Tun5}, and LSP^{Tun6}.

tion. An ESX-5 mutant of *Mycobacterium marinum* was found to affect bacillus spread to uninfected macrophages (1), and a homologue of the PPE25 gene in *M. avium* appeared to affect growth in macrophages as well as in mice (33). To our knowledge, a deletion at this locus has never been reported hitherto, suggesting that it may be deleterious to the survival of the bacillus. The fact that LSP^{Tun2} deletion was identified in four isolates suggests that bacilli bearing this deletion are not restricted in the ability to survive and to be transmitted, which is surprising given the hypothesized role in virulence of the deleted PPE25 gene. One potential hypothesis and explanation may be the fact that the deletion did not affect the actual genes involved in the formation of the ESX-5 transporter system and that other PE and PPE duplicates from this region may be able to complement for the loss of the PPE25 gene (23).

To sum up, our microarray analysis disclosed frequent natural homologous recombination within and between PE/PPE genes. Since PE/PPE gene families have been found to be enriched in essential genes and to play critical roles in host-pathogen interactions, such a propensity for recombination could represent an ideal adaptive mechanism that ensures the creation of new recombined variant molecules in response to new selective immune pressure. Moreover, since these multigene families are scattered throughout the whole genomes of *M. tuberculosis* complex strains, they are expected to contribute to genome plasticity. Future investigation of a larger set of *M. tuberculosis* samples covering the whole spectrum of strains from all three principal genetic groups (groups 1, 2, and 3), as well as representative of the global distribution of *M. tuberculosis* strains, is necessary to examine the extent and frequency of LSPs associated with these genes on a global scale. Finally, the increased tendency of PE/PPE genes to recombine is an important observation in terms of subunit vaccine formulation, as these genes are likely to be able to change rapidly in response to selective pressure (27, 42).

ACKNOWLEDGMENTS

We are indebted to Max Arella (INRS-Institut Armand-Frappier, Montreal, Quebec, Canada) and to Luke Masson (Biotechnology Research Institute, Montreal, Quebec, Canada) for supporting this collaborative study. We acknowledge Maherzia Ben Fadhel for technical assistance with sequencing and Jose R. Lappa e Silva and Luz Lazzarini for the donation of the 318/02 (RD^{Rio}) strain. We also thank Andrea Gibson and Natalia Kurepina for DNA isolation. The excellent assistance of Liliane Meunier and Miria Elias in printing the microarrays is gratefully acknowledged.

This work received financial support from the Tunisian Ministry of Higher Education, Scientific Research and Technology and from the United Nations Development Program/World Bank/World Health Organization Special Program for Research and Training in Tropical Diseases (T.D.R.). Partial funding support was obtained from NIH grants R21 AI063147 and R21 AI063147 (to J.L.H.).

REFERENCES

1. **Abdallah, A. M., T. Verboom, F. Hannes, M. Safi, M. Strong, D. Eisenberg, R. J. P. Musters, C. M. J. E. Vandenbroucke-Grauls, J. Ben Appelmelk, J. Luirink, and W. Bitter.** 2006. A specific secretion system mediates PPE41 transport in pathogenic mycobacteria. Mol. Microbiol. **62:**667–679.
2. **Aldovini, A., R. N. Husson, and R. A. Young.** 1993. The *uraA* locus and homologous recombination in *Mycobacterium bovis* BCG. J. Bacteriol. **175:** 7282–7289.
3. **Balsubramanian, V., M. S. J. Pavelka, S. S. Bardarov, J. Martin, T. R. Weisbrod, R. A. McAdam, B. R. Bloom, and W. R. Jacobs, Jr.** 1996. Allelic exchange in *Mycobacterium tuberculosis* with long linear recombination substrates. J. Bacteriol. **178:**273–279.
4. **Banu, S., N. Honore, B. Saint-Joanis, D. Philpott, M. C. Prevost, and S. T. Cole.** 2002. Are the PE-PGRS proteins of *Mycobacterium tuberculosis* variable surface antigens? Mol. Microbiol. **44:**9–19.
5. **Baulard, A., L. Kremer, and C. Locht.** 1996. Efficient homologous recombination in fast-growing and slow-growing mycobacteria. J. Bacteriol. **178:** 3091–3098.
6. **Behr, M. A., M. A. Wilson, W. P. Gill, H. Salamon, G. K. Schoolnik, S. Rane, and P. M. Small.** 1999. Comparative genomics of BCG vaccines by whole genome DNA microarray. Science **284:**1520–1523.
7. **Boshoff, H. I., M. B. Reed, C. E. Barry, and V. Mizrahi.** 2003. DnaE2 polymerase contributes to *in vivo* survival and the emergence of drug resistance in *Mycobacterium tuberculosis*. Cell **113:**183–193.
8. **Brennan, M. J., G. Delogu, Y. Chen, S. Bardarov, J. Kriakov, M. Alavi, and W. R. Jacobs, Jr.** 2001. Evidence that mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. Infect. Immun. **69:**7326–7333.
9. **Brennan, M. J., and G. Delogu.** 2002. The PE multigene family: a 'molecular mantra' for mycobacteria. Trends Microbiol. **10:**246–249.
10. **Brosch, R., S. V. Gordon, M. Marmiesse, P. Brodin, C. Buchrieser, K. Eiglmeier, T. Garnier, C. Gutierrez, G. Hewinson, K. Kremer, L. M. Parsons, A. S. Pym, S. Samper, D. van Soolingen, and S. T. Cole.** 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc. Natl. Acad. Sci. USA **99:**3684–3689.
11. **Brosch, R., S. V. Gordon, T. Garnier, K. Eiglmeier, W. Frigui, P. Valenti, S. Dos Santos, S. Duthoy, C. Lacroix, C. Garcia-Pelayo, J. K. Inwald, P. Golby, J. N. Garcia, R. G. Hewinson, M. A. Behr, M. A. Quail, C. Churcher, B. G. Barrell, J. Parkhill, and S. T. Cole.** 2007. Genome plasticity of BCG and impact on vaccine efficacy. Proc. Natl. Acad. Sci. USA **104:**5596–5601.
12. **Brudey, K., J. R. Driscoll, L. Rigouts, W. M. Prodinger, A. Gori, S. A. Al-Hajoj, C. Allix, L. Aristimuño, J. Arora, V. Baumanis, L. Binder, P. Cafrune, A. Cataldi, S. Cheong, R. Diel, C. Ellermeier, J. T. Evans, M. Fauville-Dufaux, S. Ferdinand, D. G. de Viedma, C. Garzelli, L. Gazzola, H. M. Gomes, M. C. Guttierez, P. M. Hawkey, P. D. van Helden, G. V. Kadival, B. N. Kreiswirth, K. Kremer, M. Kubin, S. P. Kulkarni, B. Liens, T. Lillebaek, M. L. Ho, C. Martin, C. Martin, I. Mokrousov, O. Narvskaïa, Y. F. Ngeow, L. Naumann, S. Niemann, I. Parwati, Z. Rahim, V. Rasolofo-Razanamparany, T. Rasolonavalona, M. L. Rossetti, S. Rüsch-Gerdes, A. Sajduda, S. Samper, I. G. Shemyakin, U. B. Singh, A. Somoskovi, R. A. Skuce, D. van Soolingen, E. M. Streicher, P. N. Suffys, E. Tortoli, T. Tracevska, V. Vincent, T. C. Victor, R. M. Warren, S. F. Yap, K. Zaman, F. Portaels, N. Rastogi, and C. Sola.** 2006. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol. **6:**23.
13. **Chaitra, M. G., S. Hariharaputran, N. R. Chandra, M. S. Shaila, and R. Nayak.** 2005. Defining putative T cell epitopes from PE and PPE families of proteins of *Mycobacterium tuberculosis* with vaccine potential. Vaccine **23:** 1265–1272.
14. **Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, and B. G. Barrell.** 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature **393:**537–544.
15. **Delogu, G., and M. J. Brennan.** 2001. Comparative immune response to PE and PE_PGRS antigens of *Mycobacterium tuberculosis*. Infect. Immun. **69:** 5606–5611.
16. **Delogu, G., C. Pusceddu, A. Bua, G. Fadda, M. J. Brennan, and S. Zanetti.** 2004. Rv1818c-encoded PE_PGRS protein of *Mycobacterium tuberculosis* is surface exposed and influences bacterial cell structure. Mol. Microbiol. **52:** 725–733.
17. **Dheenadhayalan, V., G. Delogu, and M. J. Brennan.** 2006. Expression of the PE_PGRS 33 protein in *Mycobacterium smegmatis* triggers necrosis in macrophages and enhanced mycobacterial survival. Microbes Infect. **8:**262–272.
18. **Espitia, C., J. P. Laclette, M. Mondragón-Palomino, A. Amador, J. Campuzano, A. Martens, M. Singh, R. Cicero, Y. Zhang, and C. Moreno.** 1999. The PE-PGRS glycine-rich proteins of *Mycobacterium tuberculosis*: a new family of fibronectin-binding proteins? Microbiology **145:**3487–3495.
19. **Fleischmann, R. D., D. Alland, J. A. Eisen, L. Carpenter, O. White, J. Peterson, R. DeBoy, R. Dodson, M. Gwinn, D. Haft, E. Hickey, J. F. Kolonay, W. C. Nelson, L. A. Umayam, M. Ermolaeva, S. L. Salzberg, A. Delcher, T. Utterback, J. Weidman, H. Khouri, J. Gill, A. Mikula, W. Bishai, W. R. Jacobs, Jr., J. C. Venter, and C. M. Fraser.** 2002. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. J. Bacteriol. **184:**5479–5490.
20. **Gao, Q., K. E. Kripke, A. J. Saldanha, W. Yan, S. Holmes, and P. M. Small.** 2005. Gene expression diversity among *Mycobacterium tuberculosis* clinical isolates. Microbiology **151:**5–14.
21. **Garcia-Pelayo, M. C., K. C. Caimi, J. K. Inwald, J. Hinds, F. Bigi, M. I.**

Romano, D. van Soolingen, R. G. Hewinson, A. Cataldi, and S. V. Gordon. 2004. Microarray analysis of *Mycobacterium microti* reveals deletion of genes encoding PE-PPE proteins and ESAT-6 family antigens. Tuberculosis **84:** 159–166.

22. **Garnier, T., K. Eiglmeier, J. C. Camus, N. Medina, H. Mansoor, M. Pryor, S. Duthoy, S. Grondin, C. Lacroix, C. Monsempe, S. Simon, B. Harris, R. Atkin, J. Doggett, R. Mayes, L. Keating, P. R. Wheeler, J. Parkhill, B. G. Barrell, S. T. Cole, S. V. Gordon, and R. G. Hewinson.** 2003. The complete genome sequence of *Mycobacterium bovis*. Proc. Natl. Acad. Sci. USA **100:** 7877–7882.
23. **Gey van Pittius, N. C., S. L. Sampson, H. Lee, Y. Kim, P. D. van Helden, and R. Warren.** 2006. Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (*esx*) gene cluster regions. BMC Evol. Biol. **6:**95.
24. **Gibson, A. L., R. C. Huard, N. C. Gey van Pittius, L. C. Lazzarini, J. Driscoll, N. Kurepina, T. Zozio, C. Sola, S. M. Spindola, A. L. Kritski, D. Fitzgerald, K. Kremer, H. Mardassi, P. Chitale, J. Brinkworth, D. Garcia de Viedma, B. Gicquel, J. W. Pape, D. van Soolingen, B. N. Kreiswirth, R. M. Warren, P. D. van Helden, N. Rastogi, P. N. Suffys, J. Lapa e Silva, and J. L. Ho.** 2008. Application of sensitive and specific molecular methods to uncover global dissemination of the major RDRio sublineage of the Latin American-Mediterranean *Mycobacterium tuberculosis* spoligotype family. J. Clin. Microbiol. **46:**1259–1267.
25. **Gordon, S. V., K. Eiglmeier, T. Garnier, R. Brosch, J. Parkhill, B. Barrell, S. T. Cole, and R. G. Hewinson.** 2001. Genomics of *Mycobacterium bovis*. Tuberculosis **81:**157–163.
26. **Gutacker, M. M., B. Mathema, H. Soini, E. Shashkina, B. N. Kreiswirth, E. A. Graviss, and J. M. Musser.** 2006. Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. J. Infect. Dis. **193:**121–128.
27. **Hebert, A. M., S. Talarico, D. Yang, R. Durmaz, C. F. Marrs, L. Zhang, B. Foxman, and Z. Yang.** 2007. DNA polymorphisms in the *pepA* and PPE18 genes among clinical strains of *Mycobacterium tuberculosis*: implications for vaccine efficacy. Infect. Immun. **75:**5798–5805.
28. **Karboul, A., N. C. Gey van Pittius, A. Namouchi, V. Vincent, C. Sola, N. Rastogi, P. Suffys, M. Fabre, A. Cataldi, R. C. Huard, N. Kurepina, B. Kreiswirth, J. L. Ho, M. C. Gutierrez, and H. Mardassi.** 2006. Insights into the evolutionary history of tubercle bacilli as disclosed by genetic rearrangements within a PE_PGRS duplicated gene pair. BMC Evol. Biol. **6:**107.
29. **Kato-Maeda, M., J. T. Rhee, T. R. Gingeras, H. Salamon, J. Drenkow, N. Smittipat, and P. M. Small.** 2001. Comparing genomes within the species *Mycobacterium tuberculosis*. Genome Res. **11:**547–554.
30. **Krzywinska, E., J. Krzywinski, and J. S. Schorey.** 2004. Naturally occurring horizontal gene transfer and homologous recombination in *Mycobacterium*. Microbiology **150:**1707–1712.
31. **Lazzarini, L. C., R. C. Huard, N. L. Boechat, H. M. Gomes, M. C. Oelmann, N. Kurepina, N. Shashkina, F. C. Q. Mello, A. L. Gibson, M. J. Virginio, A. G. Marsico, W. R. Butler, B. N. Kreiswirth, P. Suffys, J. R. L. Silva, and J. L. Ho.** 2007. Discovery of a novel *Mycobacterium tuberculosis* lineage that is a major cause of tuberculosis in Rio De Janeiro, Brazil. J. Clin. Microbiol. **45:**3891–3902.
32. **Lazzarini, L. C., S. M. Spindola, H. Bang, A. L. Gibson, S. Weisenberg, W. da Silva Carvalho, C. J. Augusto, R. C. Huard, A. L. Kritski, and J. L. Ho.** 2008. RDRio *Mycobacterium tuberculosis* infection is associated with a higher frequency of cavitary pulmonary disease. J. Clin. Microbiol. **46:**2175–2183.
33. **Li, Y., E. Miltner, M. Wu, M. Petrofsky, and L. E. Bermudez.** 2005. A *Mycobacterium avium* PPE gene is associated with the ability of the bacterium to grow in macrophages and virulence in mice. Cell. Microbiol. **7:**539–548.
34. **Liu, X., M. M. Gutacker, J. M. Musser, and Y. X. Fu.** 2006. Evidence for recombination in *Mycobacterium tuberculosis*. J. Bacteriol. **188:**8169–8177.
35. **Manabe, Y. C., A. M. Dannenberg, Jr., S. K. Tyagi, C. L. Hatem, M. Yoder, S. C. Woolwine, B. C. Zook, M. L. M. Pitt, and W. R. Bishai.** 2003. Different strains of *Mycobacterium tuberculosis* cause various spectrums of disease in the rabbit model of tuberculosis. Infect. Immun. **71:**6004–6011.
36. **Marmiesse, M., P. Brodin, C. Buchrieser, C. Gutierrez, N. Simoes, V. Vincent, P. Glaser, S. T. Cole, and R. Brosch.** 2004. Macro-array and bioinformatic analyses reveal mycobacterial core genes, variation in the ESAT-6 gene family and new phylogenetic markers for the *Mycobacterium tuberculosis* complex. Microbiology **150:**483–496.
37. **Namouchi, A., A. Karboul, B. Mhenni, N. Khabouchi, R. Haltiti, R. Ben Hassine, B. Louzir, A. Chabbou, and H. Mardassi.** 2008. Genetic profiling of *Mycobacterium tuberculosis* in Tunisia: predominance and evidence for the establishment of a few genotypes. J. Med. Microbiol. **57:**864–872.
38. **Narayanan, S., S. Gagneux, L. Hari, A. G. Tsolaki, S. Rajasekhar, P. R. Narayanan, P. M. Small, S. Holmes, and K. DeRiemer.** 2007. Genomic interrogation of ancestral *Mycobacterium tuberculosis* from south India. Infect. Genet. Evol. **8:**474–483.
39. **Palaci, M., R. Dietze, D. J. Hadad, F. K. Ribeiro, R. L. Peres, S. A. Vinhas, E. L. Maciel, V. do Valle Dettoni, L. Horter, W. H. Boom, J. L. Johnson, and K. D. Eisenach.** 2007. Cavitary disease and quantitative sputum bacillary load in cases of pulmonary tuberculosis. J. Clin. Microbiol. **45:**4064–4066.
40. **Ramakrishnan, L., N. A. Federspiel, and S. Falkow.** 2000. Granuloma-specific expression of Mycobacterium virulence proteins from the glycine-rich PE-PGRS family. Science **288:**1436–1439.
41. **Singh, K. K., X. Zhang, A. S. Patibandla, P. Chien, Jr., and S. Laal.** 2001. Antigens of *Mycobacterium tuberculosis* expressed during preclinical tuberculosis: serological immunodominance of proteins with repetitive amino acid sequences. Infect. Immun. **69:**4185–4191.
42. **Skeiky, Y. A., and J. C. Sadoff.** 2006. Advances in tuberculosis vaccine strategies. Nat. Rev. Microbiol. **6:**469–476.
43. **Stewart, G. R., J. Patel, B. D. Robertson, A. Rae, and D. B. Young.** 2005. Mycobacterial mutants with defective control of phagosomal acidification. PLoS Pathog. **1:**269–278.
44. **Strong, M., M. R. Sawaya, S. Wang, M. Phillips, D. Cascio, and D. Eisenberg.** 2006. Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. USA **103:**8060–8065.
45. **Talaat, A. M., R. Lyons, S. T. Howard, and S. A. Johnston.** 2004. The temporal expression profile of *Mycobacterium tuberculosis* infection in mice. Proc. Natl. Acad. Sci. USA **101:**4602–4607.
46. **Talarico, S., M. D. Cave, C. F. Marrs, B. Foxman, L. Zhang, and Z. Yang.** 2005. Variation of the *Mycobacterium tuberculosis* PE_PGRS33 gene among clinical isolates. J. Clin. Microbiol. **43:**4954–4960.
47. **Talarico, S., L. Zhang, C. F. Marrs, B. Foxman, M. D. Cave, M. J. Brennan, and Z. Yang.** 2008. *Mycobacterium tuberculosis* PE_PGRS16 and PE_PGRS26 genetic polymorphism among clinical isolates. Tuberculosis (Edinburgh) **88:**283–294.
48. **Tsolaki, A. G., A. E. Hirsh, K. DeRiemer, J. A. Enciso, M. Z. Wong, M. Hannan, Y. O. Goguet de la Salmoniere, K. Aman, M. Kato-Maeda, and P. M. Small.** 2004. Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. Proc. Natl. Acad. Sci. USA **101:**4865–4870.
49. **Tundup, S., Y. Akhter, D. Thiagarajan, and S. E. Hasnain.** 2006. Clusters of PE and PPE genes of *Mycobacterium tuberculosis* are organized in operons: evidence that PE Rv2431c is co-transcribed with PPE Rv2430c and their gene products interact with each other. FEBS Lett. **580:**1285–1293.
50. **van Soolingen, D., P. W. Hermans, P. E. de Haas, D. R Soll, and J. D. van Embden.** 1991. Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. J. Clin. Microbiol. **29:**2578–2586.
51. **Voskuil, M. I., D. Schnappinger, R. Rutherford, Y. Liu, and G. K. Schoolnik.** 2004. Regulation of the *Mycobacterium tuberculosis* PE/PPE genes. Tuberculosis (Edinburgh) **84:**256–262.
52. **Wang, J., Y. Qie, H. Zhang, B. Zhu, Y. Xu, W. Liu, J. Chen, and H. Wang.** 2008. PPE protein (Rv3425) from DNA segment RD11 of *Mycobacterium tuberculosis*: a novel immunodominant antigen of *Mycobacterium tuberculosis* induces humoral and cellular immune responses in mice. Microbiol. Immunol. **52:**224–230.
53. **Wang, X., and B. Seed.** 2003. Selection of oligonucleotide probes for protein coding sequences. Bioinformatics **19:**796–802.
54. **Zhang, H., J. Wang, J. Lei, M. Zhang, Y. Yang, Y. Chen, and H. Wang.** 2007. PPE protein (Rv3425) from DNA segment RD11 of *Mycobacterium tuberculosis*: a potential B-cell antigen used for serological diagnosis to distinguish vaccinated controls from tuberculosis patients. Clin. Microbiol. Infect. **13:** 139–145.